

AD-A118 447

NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/6 5/9
COMPUTERIZED ADAPTIVE TESTING PROJECT: OBJECTIVES AND REQUIREME--ETC(U)
JUL 82 J R MCBRIDE

UNCLASSIFIED

NPRDC-TN-82-22

NL

1-1
1-1



END
DATE
FILMED
9 '82
DTIC

12

NPRDC TN 82-22

JULY 1982

AD A118447

**COMPUTERIZED ADAPTIVE TESTING
PROJECT: OBJECTIVES AND
REQUIREMENTS**

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED



DTIC
ELECTE
AUG 23 1982
S **D**
H

**NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152**

DTIC FILE COPY

82 08 23 050

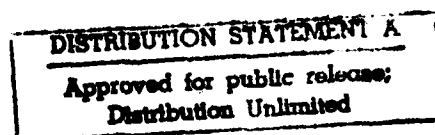


**COMPUTERIZED ADAPTIVE TESTING PROJECT:
OBJECTIVES AND REQUIREMENTS**

James R. McBride

Reviewed by
Martin F. Wiskoff

Released by
James F. Kelly, Jr.
Commanding Officer



Navy Personnel Research and Development Center
San Diego, California 92152

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPRDC TN 82- 22	2. GOVT ACCESSION NO. ADA118 447	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPUTERIZED ADAPTIVE TESTING PROJECT: OBJECTIVES AND REQUIREMENTS		5. TYPE OF REPORT & PERIOD COVERED Technical Note
7. AUTHOR(s) James R. McBride		6. PERFORMING ORG. REPORT NUMBER 12-82-04
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		8. CONTRACT OR GRANT NUMBER(s)
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Z1385-PN
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE July 1982
		13. NUMBER OF PAGES 13
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Computerized adaptive testing Psychological testing Tailored testing Personnel testing Computer systems design ASVAB (Armed Services Vocational Aptitude Battery)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) As lead laboratory in a joint-service project, NAVPERSRANDCEN is contracting for design, development, testing, and evaluation of a system for automated adaptive administration of military personnel selection tests. This report describes the planned contracting approach and system requirements for developing a computerized adaptive testing system (CAT). A		

DD FORM 1 JAN 79 1473

EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-LF-014-6401

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

S/N 0102- LP-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FOREWORD

A joint-service coordinated effort is in progress to develop a computerized adaptive testing (CAT) system and to evaluate its potential for use in the Military Enlistment Processing Stations as a replacement for the Armed Services Vocational Aptitude Battery (ASVAB). The Navy Personnel Research and Development Center has been designated lead laboratory for this effort.

This report is intended to serve as a working paper documenting CAT system functional requirements and schedules. It was developed within engineering development project Z1385-PN (Computerized Adaptive Testing), under the mission sponsorship of the Chief of Naval Operations (OP-115), and is intended for the ASVAB Executive Steering Committee, the CAT Interservice Coordinating Committee, and the ASVAB Working Group. A more detailed technical discussion of functional requirements for a CAT system will be provided in subsequent reports.

JAMES F. KELLY, JR.
Commanding Officer

JAMES J. REGAN
Technical Director



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
(Avail and/or	
Dist	Special
A	

CONTENTS

	Page
INTRODUCTION	1
Problem and Background	1
Objectives	2
APPROACH	2
SYSTEM REQUIREMENTS	4
Application	4
Setting	4
Performance	5
Reliability	5
Security	5
Flexibility	5
Expandability	6
Cost	6
REFERENCES	7
DISTRIBUTION LIST	9

INTRODUCTION

Problem and Background

NAVPERSRANDCEN is contracting for design, development, testing, and evaluation of a system for automated adaptive administration of military personnel selection tests. A brief discussion is needed to document the system's rationale, functional requirements, and proposed schedules.

The system will be evaluated for its potential to replace the current paper-and-pencil Armed Service Vocational Aptitude Battery (ASVAB), which is used by all four armed services for enlisted personnel selection and classification. There are a number of deficiencies inherent in the paper-and-pencil ASVAB, such as:

1. Excessive duration of personnel test sessions.
2. Poor measurement precision at high and low ability levels.
3. Susceptibility to theft, compromise, and coaching.
4. Expense of printing, distribution, and storage of multiple forms of test booklets and answer sheets.
5. Susceptibility to errors inherent in manual score tallying, score conversion, computation of score composites, and score recording.
6. Long lead time and high expense required to develop replacement test forms.

The automated adaptive system is intended as a vehicle for applying a psychometric technology called computerized adaptive testing (CAT). CAT is an emerging technology that has been the subject of extensive research dating back to the 1950s (e.g., Weiss, 1974; Lord, 1974), most of which has been sponsored by the Office of Naval Research. Exploratory and advanced development of CAT applications has been conducted at the U.S. Civil Service Commission (Clark, 1976; Urry, 1977) and, more recently, at the Educational Testing Service (Lord, 1977a, b), NAVPERSRANDCEN (McBride, 1980), the Army Research Institute (ARI) (McBride, 1979), and the Air Force Human Resources Laboratory (AFHRL) (Ree & Jensen, 1980).

In 1977, a NAVPERSRANDCEN exploratory development project, entitled USMC Computerized Adaptive Testing, was initiated in response to a Marine Corps Statement of Requirement (SOR) for research and development in the area of personnel accessions systems. A focal issue of that SOR was the desirability of evaluating computer-based adaptive personnel tests as potential alternatives to paper-and-pencil tests. CAT was viewed as a single solution to the many problems inherent in the ASVAB testing program. Enthusiasm for the concept of computer-administered adaptive tests was communicated by the Marine Corps Headquarters to policy personnel on the staffs of the Assistant Secretaries (for Manpower, Research Affairs and Logistics) of the Navy and of Defense. In January 1979, a joint-service effort was initiated (1) to evaluate the feasibility of implementing CAT in the Department of Defense and (2) to perform the additional research and development needed. The Navy was designated the lead service for this effort. An interservice coordinating committee was created, with the Navy responsible for chairmanship. Subsequently, NAVPERSRANDCEN was designated lead laboratory for the project.

Objectives

The goal of the joint-service project is to design, develop, test, and evaluate a system for automated, adaptive administration of the armed services' personnel selection and classification tests. The technical feasibility of CAT has been demonstrated in previous exploratory development. The primary objective of this phase of the project was to develop a cost-effective CAT system suitable for nationwide implementation in the Military Enlistment Processing Stations (MEPS) (formerly called Armed Forces Examining and Entrance Stations (AFEES)). The system will be designed to replace the paper-and-pencil ASVAB.

APPROACH

In producing the CAT system, two avenues of approach are required: psychometric construction of the adaptive testing procedures and tests, and engineering development of the system to implement those procedures and tests. NAVPERSRANDCEN is responsible for developing the psychometric procedures and AFHRL, for constructing the tests. As lead laboratory, NAVPERSRANDCEN is contracting for the engineering development work. The planned contracting approach is to sponsor industry competition to design and produce a feasible, cost-effective CAT system capable of incorporating the following characteristics:

1. On-line interactive administration of personnel tests, using automated display and response media.
2. Dynamic tailoring of test difficulty to each examinee's ability, contingent on performance at earlier stages of the test (Lord, 1977a; Owen, 1975; Samejima, 1977; Urry, 1977). This tailoring will reduce ASVAB test duration by 50 percent or more and will improve measurement precision at the extremes of ability levels.
3. Administration of a unique tailored sequence of test items to each examinee. Each test is drawn from a very large bank of test items. The size of the item bank will effect a significant defense against test compromise and coaching.
4. Replacement of all printed test material by electronic media, thus eliminating printing and storage costs of test booklets and answer sheets. This replacement will also enhance test security because no printed materials will be available for theft.
5. Computerized scoring, score conversion (Bejar & Weiss, 1979), score composite computation, and score recording, thus eliminating erroneous personnel record data attributable to clerical errors in manual scoring and recording.
6. On-line administration of experimental replacement test questions to examinees, thus reducing the lead time needed to construct replacement test forms.

The engineering development part of this program is divided into three stages: (1) system design, prototype production, and demonstration competition, (2) full-scale production, model field testing, and evaluation, and (3) operational implementation. Stage 1 involves two or more parallel contractual efforts to design cost-effective versions of a CAT system; that is, its contracts contain options to produce and demonstrate working CAT prototypes and to submit system implementation proposals. Prototypes and proposals obtained will result in a Stage 2 production unit, which will be subjected to field testing and evaluation. If the evaluation is positive and if cost-benefit data warrant,

production and implementation of a nationwide CAT system are expected to follow. Table 1 presents a timetable for CAT system development. The target date for initial operational capability (IOC) of the contemplated CAT system is December 1984.

Table 1
Timetable for the Three-stage Engineering Development
Of the Computerized Adaptive Testing System

Event	Stage		
	1	2	3
Request for proposals (RFP) issue	Jul 1981	Jan 1983	Jan 1984
Proposal deadline	Sep 1981	Apr 1983	Mar 1984
Contract award(s)	Dec 1981	Jul 1983	Jun 1984
Effort completed	Jun 1983	Dec 1983	+
Duration ^a	18 months	6 months	+

^aDuration of operational implementation will be negotiated during the Stage 3 contracting process.

The selection of contractors at Stage 1 was fully competitive. Competition for Stage 2 and 3 awards will be limited to the contractors at preceding stages. The number of Stage 1 contracts awarded was determined by the quality of the proposals, the amount of funds available, and the costs proposed by the separate offerors.

Because each Stage 1 contractor is a potential prime contractor for nationwide CAT system implementation, the Stage 1 competition was limited to firms with the capability to perform system design and prototype development. Factors also considered included the corporate size, capital, and experience to perform component fabrication, system deployment and installation, system maintenance, and training in system operations and maintenance. In addition, because of the highly specialized nature of the psychometric technology underlying CAT, offerors must identify key personnel with appropriate psychometric training and experience who will be assigned to the CAT development effort.

The materials and services entailed in the operational implementation stage will require substantial initial investment and an extended period of preparation by the contractor. For this reason, it is intended that a Stage 2 contractor be awarded the prime contract for operational implementation (subject to successful field testing and evaluation, as well as to decision by appropriate authority). The scope and magnitude, as well as the potential nonmilitary applications of the Stage 3 contract, should serve as incentive for the best efforts of the competitors at Stages 1 and 2. Accordingly, cost-sharing contracts are being employed during Stage 1.

SYSTEM REQUIREMENTS

Application

The CAT system is intended to replace the paper-and-pencil test battery (ASVAB) now used for armed services enlisted personnel selection and classification. As a replacement for ASVAB, the CAT system must provide a counterpart for every critical aspect and function of the paper-and-pencil tests, including:

1. Adaptive counterparts of each ASVAB subtest.
2. A permanent record of test scores.
3. A detailed record of each test administered.
4. An interface with the AFEEES reporting system (ARS).
5. Capability to administer experimental tests.
6. Provision for replacement and updating of test questions.
7. Provision for identifying likely instances of test compromise and malingering.

Setting

The operational ASVAB program is administered by the Military Enlistment Processing Command (MEPCOM). MEPCOM operates 69 permanent examining sites called MEPS (Military Enlistment Processing Stations). Each MEPS controls one or more remote operations called Mobile Examining Team (MET) sites, the sole function of which is ASVAB test administration to applicants for military enlistment. In 1979, MEPCOM administered ASVAB tests to about 650,000 applicants in the MEPS/MET system. Since a number of these applicants were tested more than once, approximately 800,000 ASVAB tests were given.

About 70 percent of all ASVAB testing takes place at MET sites and the other 30 percent, at MEPS. Each MEPS is a permanent installation, dedicated to enlistment examining and processing, each having storage space, a testing room, and staff dedicated to personnel testing. In contrast, MET sites typically are not permanent installations. MET site testing typically takes place in "borrowed" rooms and is typically administered by part-time personnel, either military personnel assigned to MEPCOM or civilian personnel employed by the Office of Personnel Management (OPM). MEPCOM contracts with OPM for administration of ASVAB in selected MET sites.

MET sites typically have no provisions for storage of test materials. Instead, the examiner carries the tests, forms, and answer sheets to and from the MET sites for each session. All operational ASVAB tests are scored at the MEPS. After preliminary scoring of a portion of each test, the MET site examiner forwards the ASVAB answer sheet to the MEPS for scoring. Expeditious means, such as express mail or Federal Express, are often used.

Detailed descriptions of operations at MEPS and MET sites are contained in Section 1 of a report on CAT prepared by the Rehab Group (1980a). The differences between the

two types of site may have implications for the design of a CAT system intended to accommodate both types. At the system design stage, contractors will be required to address separately the feasibility of their designs for MEPS and for MET sites. Different designs may be appropriate or necessary.

The Rehab Group report includes a discussion of the volume of testing as it varies from one MEPS to another and from month to month. The CAT system must be designed not only to accommodate the normal volume of testing, but also to handle the volume planned in the event of mobilization. Such volume may require additional testing shifts or rapid system expansion.

Performance

The Rehab Group report indicates that it requires more than 4 hours to prepare and administer ASVAB forms; testing alone takes about 2-3/4 hours. One goal of the CAT project is to reduce testing time by half or more. This goal is feasible because of the adaptive nature of CAT: Fewer questions are needed to achieve the measurement precision of a counterpart paper-and-pencil test. However, the efficiency afforded by adaptive testing may be offset if system response time at the display terminal is slow. Therefore, the system must be designed so that the time interval between examinee response and display of the next question is very short: nominally, 1 second or less on the average, with perhaps a 3-second upper limit. During this interval, the system must process the examinee's response, select the next question, retrieve it, and display it. Some procedures for adaptive testing entail intensive numerical computation to optimize selection of the next item.

Reliability

Because of the nature of enlistment processing, it is imperative that processing be completed for every applicant who arrives. Furthermore, it is both expensive and undesirable for processing to take more than 1 day. Thus, the test administration components of the CAT system at each testing site must be characterized by extremely high reliability—virtually as high as the reliability of paper-and-pencil tests. By the same token, the system must have availability and maintainability characteristics sufficient to achieve the same goal: meeting each day's testing schedule.

Security

Experience indicates that there will be efforts to compromise the integrity of the selection tests. In the past, these efforts have included theft of test booklets, cheating, coaching of examinees, and possible impersonation of applicants and falsification of records. Both hardware and software must be designed to defeat sophisticated attempts to violate the security of the test questions and the test-scoring and score-recording subsystems.

Flexibility

One important advantage of automated tests over ASVAB is the potential for routine administration of experimental test questions to all examinees. This capability would ameliorate the present cumbersome requirement for the large-scale, extra test administration sessions that collect data needed for test construction, item analysis, and norming of new ASVAB test forms. Using this capability would materially reduce the cost of developing new tests, test items, and procedures. The CAT system must be designed so that (1) potential test materials can be routinely, reliably, and inexpensively distributed to

CAT sites and administered there, and (2) the collected data can be returned by similar means to a central facility. Similar provisions for updating system software are highly desirable.

Expandability

On initial implementation, the CAT system will entail automated, adaptive administration of traditional, multiple-choice test questions. However, it is anticipated that test stimuli considerably different from those used in the current printed medium, as well as response modes other than multiple-choice, will be available in the future. In anticipation of expanding the capabilities of the CAT system, the system must be designed and engineered in a manner that will permit convenient and inexpensive upgrading of hardware and software.

Cost

The decision to implement a nationwide CAT system will be influenced by results of cost-benefit analyses. There are numerous potential benefits of CAT, many of which are tangible and involve elimination of costs inherent in the current paper-and-pencil testing system. These savings may be offset by hardware and life-cycle costs peculiar to the CAT system. A crucial aspect of CAT system feasibility will be its life-cycle cost. The cost of the ASVAB program, assessed in 1980 (Rehab Group, 1980b), should serve as a cost target for CAT. From the outset, prospective offerors and contractors should strive to develop system design concepts that will keep hardware, operating, and maintenance costs at the minimum level consistent with performance and reliability.

REFERENCES

- Bejar, I. I., & Weiss, D. J. Computer programs for scoring test data with item characteristic curve models (Res. Rep. 79-1). Minneapolis: Psychometric Methods Program, University of Minnesota, February 1979. (NTIS No. AD-A067928)
- Clark, C. L. (Ed). Proceedings of the First Conference on Computerized Adaptive Testing, June 1975. PS-75-6, Personnel Research and Development Center, U.S. Civil Service Commission. Washington, DC: U.S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-00940-9)
- Lord, F. M. Individualized testing and item characteristic curve theory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes, Contemporary developments in mathematical psychology (Vol. II). San Francisco: W. H. Freeman, 1974.
- Lord, F. M. A broad-range test of verbal ability. Applied Psychological Measurement, 1977, 1, 95-100. (a)
- Lord, F. M. Practical applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14, 117-138. (b)
- McBride, J. R. Adaptive mental testing: The state of the art (Tech. Rep. 423). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, November 1979.
- McBride, J. R. Adaptive verbal ability testing in a military setting. Proceedings of the 1979 Computerized Adaptive Testing Conference. Minnesota: Psychometric Methods Program, University of Minnesota, September 1980.
- Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70(350), 351-356.
- Ree, M. J., & Jensen, H. E. Effects of sample size on linear equating of item characteristic curve parameters. Proceedings of the 1979 Computerized Adaptive Testing Conference. Minneapolis: Psychometric Methods Program, University of Minnesota, September 1980.
- Rehab Group, Inc. Final Report: Computerized adaptive testing, assessment of requirements. Falls Church, VA: Author, July 1980. (a)
- Rehab Group, Inc. Cost target for a computerized adaptive testing (CAT) system. Falls Church, VA: Author, 1980. (b)
- Samejima, F. A use of the information function in tailored testing. Applied Psychological Measurement, 1977, 1, 233-247.
- Urry, V. W. Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 1977, 14, 181-196.
- Weiss, D. J. Strategies of adaptive ability measurement (Res. Rep. 74-6). Minneapolis: Psychometric Methods Program University of Minnesota, December 1974. (NTIS No. AD-004270)

DISTRIBUTION LIST

Chief of Naval Material (NMAT-05)

Commander, Air Force Human Resources Laboratory , Brooks Air Force Base (Scientific and Technical Information Office)

Commander, Army Research Institute for the Behavioral and Social Sciences (PERI-ASL)

Defense Technical Information Center (DDA) (12)

LMEI
-8